Identification of allele specific function of *Candida albicans* Pseudouridine Synthase 4

An Honors Thesis (HONR 499)

By

Caleb Embree

Thesis Advisor

Dr. Douglas Bernstein

Ball State University

Muncie, Indiana

April 2018

Expected Date of Graduation

May 2018

Abstract

Candida albicans is the most prevalent human fungal pathogen. Unfortunately, there still remains much that is unknown about it, including the reason for maintaining a diploid genome and the function of many genes. In this thesis I describe the phenotypic consequences of up and downregulating Pseudouridine Synthase 4 (*PUS4*) of *C. albicans,* as well as an examination of the differences in the expression level of the entire transcriptome. These experiments and resulting bioinformatic analysis have lead to our conclusion that that *PUS4A* is more lowly transcribed than *PUS4B*. This data suggests *PUS4A* may be nonfunctional or play a different role in *C. albicans* biology.

Acknowledgments

I would like to thank Dr. Bernstein for his support and training for the three years that I have been in his lab. He has guided my work throughout my undergraduate career and has helped prepare me for my future in innumerable ways.

I would like to thank Rebecca Kurtz-Paulette and Dr. Munni Begum with their help analyzing the RNA seq data, which I could not have done without them. I would also like to thank all other members of the Bernstein lab for their help throughout this project.

Process Analysis Statement

My main goal for this thesis project was to conduct basic scientific research in the hopes of answering a simple question: do the two copies of the gene for Pseudouridine Synthase 4 in the fungal pathogen *Candida albicans* result in proteins with different functions? While this research was trying to answer a relatively simple question, it was also used to examine a more important question: Why does *Candida albicans* maintain a diploid genome? Understanding the importance of having two slightly different copies of one specific gene will give us a better understanding of the reason for having two copies of every gene. To examine these questions, I created two mutant strains that allowed me to increase or decrease the expression level of either copy of the gene, depending on what type of media they were grown on. These mutants allow me to examine what changes happen to the yeast as I change the expression level of either copy of Pseudouridine Synthase 4 (*PUS4*). I also used bioinformatics tools to examine sequence differences between the alleles.

The process of conducting this thesis research, as well as the two years of research that lead up to it, taught me a lot about how to conduct scientific research. Not only did I learn how to design and conduct experiments, I also learned how to think like a scientist. This was most recently shown when our mRNA sequencing revealed new information about the expression level of expression of either allele of *Candida albicans PUS4A* and *PUS4B*. I had to take the new data and figure out how it lined up with the rest of the data that I had collected. The new data prompted me to conduct another bioinformatics analysis, which in turn revealed data that helps to make sense of all of

our previous results. If I hadn't been able to understand how the sequencing data fit with the rests of my experiments, I likely wouldn't have come to the proper conclusions.

Through the process of conducting research leading up to and during my thesis I also had to deal with the failure that is associated with conducting scientific experiments. The first year that I spent in the Bernstein lab I tried and failed to make a knockout of each allele of *PUS4* using the CRISPR/Cas9 system. I had to learn how to deal with experiments failing over and over, and come up with possible explanations for the failure, both experimentally and biologically. Along with the failures though, I also learned that sometimes the failures can be informative as well. Using the knowledge that I learned from sequencing and bioinformatics analysis I was able to draw conclusions about why my initial experiments did not work.

My results also taught me the importance of continuing to read the relevant literature. Reading more literature about *Candida albicans* was critical to putting together the information to make proper conclusions. Learning about the haploinsufficency of *PUS4* in *Candida albicans* was a major indicator that the two alleles did not have the same expression level and function, and helped to indicate that I should conduct another bioinformatics analysis. If I hadn't continued to read and learn about *Candida albicans* I likely wouldn't have had some of the breakthroughs I did.

Identification of allele specific function of *Candida albicans* Pseudouridine Synthase 4

Introduction

Candida albicans is the most prevalent human fungal pathogen, causing a variety of diseases including thrush, vaginal candidiasis, and invasive candidiasis (Kullberg and Arendrup, 2015). Invasive candidiasis alone causes 46,000 infections in the US every year, and has a mortality rate as high as 30% (CDC, 2013). As *C. albicans* strains are becoming increasingly resistant to antifungal drugs it becomes ever more important that we understand this organism at a molecular level.

Though it is a clinically relevant organism, much of the inner workings of *Candida albicans* remain a mystery. While a sequenced genome of *C. albicans* exists, many open reading frames (ORFs) are not annotated (Skrzypek et al., 2017). Additionally, many of the annotations that are present are based upon orthology to *Saccharomyces cerevisiae*. *S. cerevisiae* is a genetic relative of *C. albicans*, both are members of the order Endomycetales. However, they vary massively in key phenotypes, including environmental niche and pathogenesis (Oh et al., 2010). The differences between the two organisms makes inference of protein function through sequence analysis challenging.

Further contributing to the difficulty of understanding *Candida albicans* is that it contains a diploid genome. While it does rarely exist in a haploid state, *C. albicans* is regarded as an obligate diploid (Hickman et al., 2013). Almost all clinical isolates and laboratory strains, including the most common laboratory strain SC5314, have a diploid genome (Skrzypek et al., 2017). Furthermore, haploids quickly revert to a diploid state.

Therefore, it is important that we reach a better understanding of why *C. albicans* maintains the diploid state.

One avenue of exploration is the allelic difference that are present in many ORFs in *Candida albicans*, including Pseudouridine Synthase 4 (*PUS4)* (Skrzypek et al., 2017). Most of our knowledge of the molecular function of *C. albicans* Pseudouridine Synthase 4 comes from its S. cerevisiae ortholog. In *S. cerevisiae PUS4* encodes a pseudouridine synthase that catalyzes the formation of pseudouridine from uridine at residue 55 of tRNAs (Figure 1) (Becker et al., 1997). It operates on both cytoplasmic and mitochondrial tRNAs, as well as a less well understood role in modifying mRNA (Schwartz et al., 2014).

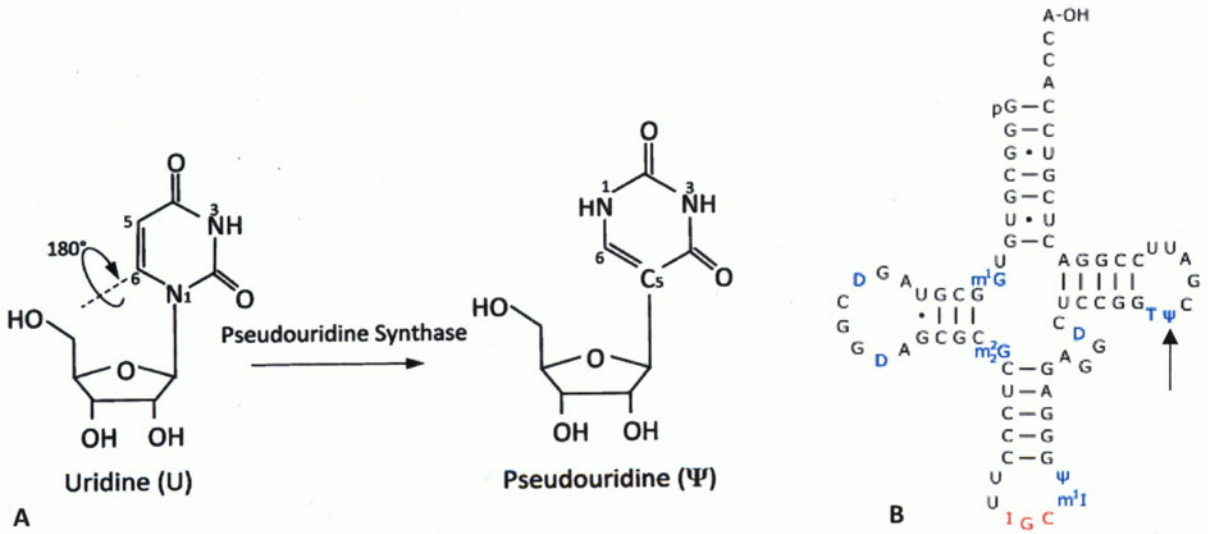

Figure 1: A) A diagram showing the process of pseudouridylation. Adapted from (Zhao et al., 2018). **B)** A sequence map of yeast tRNA showing the pseudouridine at residue 55 caused by Pus4.

Candida albicans Pus4 is believed to have the same functions as *S. cerevisiae* Pus4 (Skrzypek et al., 2017). However, the two alleles of *PUS4* have differences in sequence. The 13 nucleic acid differences results in 6 amino acid differences between the two proteins. Additionally, it is known that *PUS4* is haploinsufficent (Oh et al., 2010). These two facts lead to our hypothesis that the two alleles of *C. albicans PUS4* results in proteins with different functions. To test this I am examining the phenotypic consequences of knocking down or overexpressing either the A or B allele, as well as a examining the consequences on the *C. albicans* transcriptome during either A or B allele knockdown.

Methods

The maltose promoter

A maltose inducible promoter, pV811, was introduced immediately upstream of either the A or the B allele of *PUS4* using homologous recombination. The promoter upregulates transcription of the allele when it is grown on maltose and reduces transcription when grown on glucose. The promoter construct also conveys resistance to nourseothricin, allowing for screening of our transformed mutants. After purifying the promoter plasmid, PCR was conducted using primers with 40bp of homology to the PUS4 allele and 20bp homology to the promoter (Appendix 1). This product was transformed into *Candida albicans* and transformation success was screened using nourseothricin. The proper insertion of the promoter was verified by PCR and sequencing. Strains with properly inserted promoter were saved at -80˚C in a glycerol stock, and grown from the stock for every experiment (Appendix 2).

Oligomycin resistance and tolerance

The resistance and tolerance assay was conducted using a standard drug diffusion assay, with 100ng of Oligomycin being used on each plate. The plates were analyzed using the R package DiskImageR, and the resistance and tolerance was determined using a 50% reduction in growth (Gerstein et al., 2016).

Virulence trials

To examine *C. albicans* virulence *Galleria melonella*, wax moth larvae, were used as a model organism. For the virulence trials 1OD of yeast cells were spun at 5,000 RPM for 5 minutes in an centrifuge. The pellets were washed with 1ml PBS, and then were spun again. The pellets were suspended in 1ml PBS and 10µl ($1x10^5$ cells) were injected into the second to last proleg of a larva. The worms were placed at 37˚C and monitored for death every day for up to 12 days. The survival chart was visualized using the SurvMiner package in R.

mRNA sequencing

The total RNA was extracted by hot acid phenol from wildtype SC5314, and the *PUS4A* and *PUS4B* knockdown strains. Four replicates of each strain were polyA selected and sequenced by GeneWiz, a Next Generation Sequencing company. The reads were then mapped to the *Candida albicans* genome, and expression was compared to identify significant changes in gene expression between strains.

Results

Growth and filamentation assays

The growth rate of CE17, which contains the maltose promoter upstream of *PUS4A*, and CE16, which contains the maltose promoter upstream of *PUS4B*, were compared to wildtype *Candida albicans*, SC5314 on media containing glucose and maltose. We did not see any difference between wildtype and the knockdown strains that were grown on glucose (Figure 2). We did see that all of the yeast grown with maltose as a carbon source grew slower than the ones on glucose. However, when comparing the maltose overexpression strains to SC5314 grown on maltose, there was no difference in growth. This result suggests that Pus4 does not have a significant effect on growth.

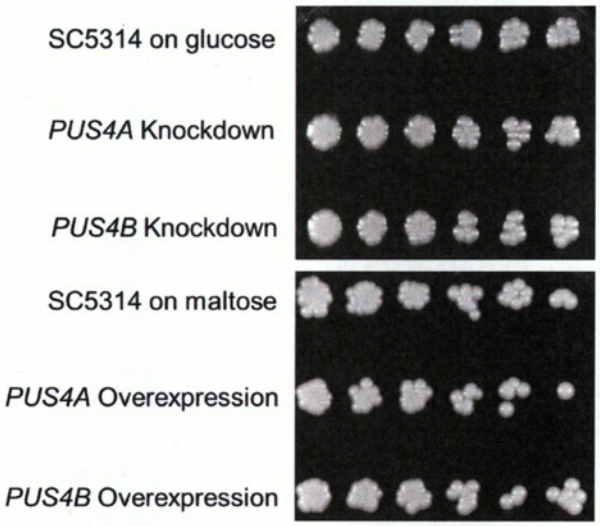

Figure 2: Growth assay conducted on YPD and YPmaltose at 30˚C, grown for 2 days. There was no difference in growth when the wildtype and strains with altered *PUS4* expression.

Oligomycin Resistance Analysis

S. cerevisiae Pus4 is critical for pseudouridylation of mitochondrial tRNAs, and the same function is inferred for *Candida albicans* Pus4. To determine how changes in *PUS4* expression affects the ability of the cell to survive with impaired mitochondria function we tested our mutants' resistance and tolerance to Oligomycin, an electron transport chain inhibitor, by disk diffusion assay. Resistance is the distance from the edge of the drug disk to where a 50% reduction in growth occurs. Tolerance is based upon the amount of cells that are growing beyond the 50% reduction in growth. There is no difference when looking at the resistance of either the knockdown or overexpression strains (Figure 3A). However, there was a reduction in tolerance when any of the yeast were grown on maltose. Furthermore, when compared to SC5314 that was grown on maltose, overexpressing both *PUS4A and PUS4B* had decreased tolerance to oligomycin. Additionally, reducing expression of *PUS4B* results in an increased tolerance to oligomycin (Figure 3B). These results suggest a possible role of Pus4 in mitochondrial function. Upregulating *PUS4B* increases tolerance to mitochondrial dysfunction, which indicates a possible change in transcription when the cell is forced to perform fermentation instead of respiration.

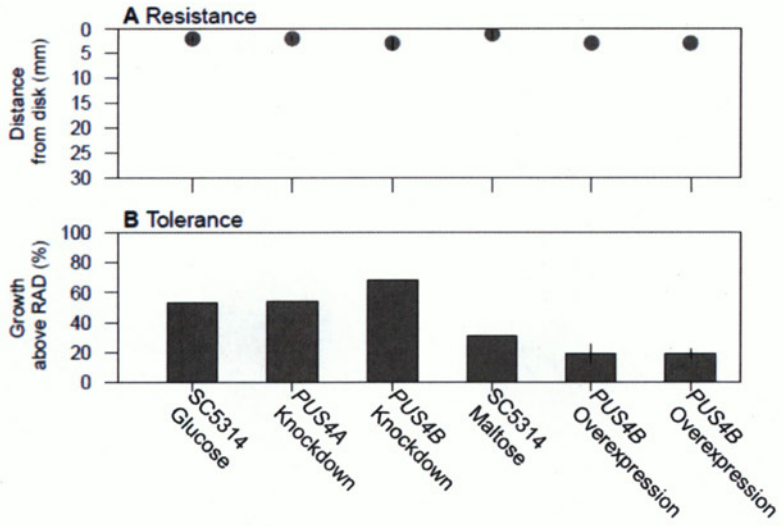

Figure 3: A) The resistance to oligomycin based on a 50% reduction in growth. There is no difference between resistance to oligomycin when the expression level of *PUS4* is changed. **B)** The tolerance to oligomycin calculated based on a 50% reduction in growth. The *PUS4B* knockdown had increased tolerance when compared to wild type, and reducing the expression of *PUS4* results in decreased tolerance to oligomycin.

Virulence Assay

As *Candida albicans* is a human pathogen, it is critical to determine if changes in *PUS4* expression results in a change in virulence. The virulence of the mutant strains was tested by injecting *Galleria melonella* with *Candida albicans*. *G. melonella* are well characterized models for *Candida albicans* virulence, and have a similar innate immune system to mammals (Brennan et al., 2002). No difference was detected between the virulence of wildtype *C. albicans* and strains with changes in *PUS4* expression (Figure 4). These results indicate that Pus4 doesn't play a role in virulence.

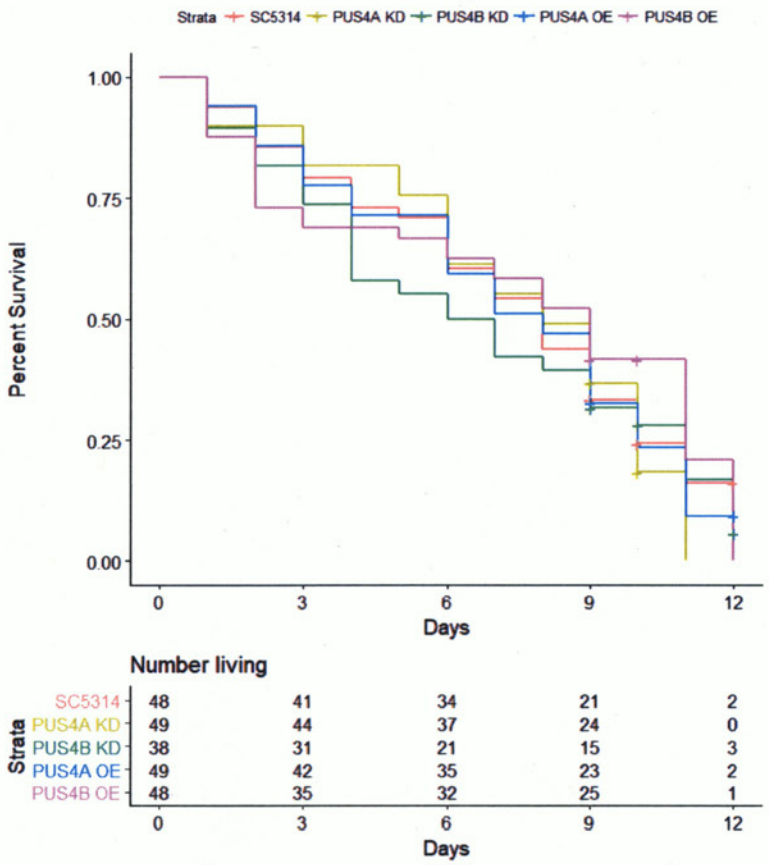

Figure 4: A survival chart showing the death of *Galleria melonella* larvae over a period of days. There was no difference in the survival of worms injected with wildtype *Candida albicans* or with strains that had changes in *PUS4* expression.

mRNA sequencing

To determine how a knockdown of the A or B allele of *PUS4* effects the expression of other genes, we sequenced the transcriptome of the two knockdown strains and wildtype *Candida albicans.* Because there are only 13 differences between the DNA sequences of *PUS4A* and *PUS4B*, most reads map nonspecifically to either allele (Skrzypek et al., 2017). Mapping reads to the genome indicates that knocking down *PUS4A* results in a negligible change in reads that map to *PUS4*. However,

knocking down *PUS4B* sees a large reduction in transcripts that map to *PUS4* (Figure 5). Our results suggest only *PUS4B* is expressed. Using a maltose promoter to knockdown *PUS4A* likely does not change the expression level because it is not expressed.

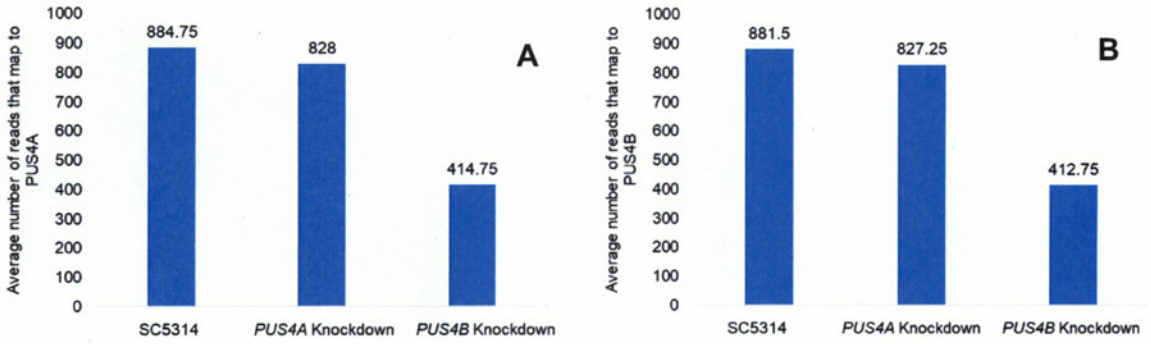

Figure 5: A) The average number of reads from RNA sequencing that mapped to *PUS4A* from the knockdown and wildtype strains. Only the *PUS4B* knockdown resulted in a significant decrease in expression. **B)** The average number of reads from RNA sequencing that map to *PUS4B* from the knockdown and wildtype strains. Only the *PUS4B* knockdown resulted in a decrease in expression.

Bioinformatic analysis

Based on the previous results we conducted a bioinformatic analysis of the sequence of Pus4. The most notable difference between the two proteins is a mutation at residue 175 of Pus4 (Skrzypek et al., 2017). In *PUS4B* there is a proline residue at this position, but in *PUS4A* it has changed to a serine. Furthermore, this proline residue is very highly conserved, appearing in the Pus4 ortholog in eight closely related *Candida* species, *Saccharomyces cerevisiae,* as well as many other yeast species that are not closely related (Figure 6). Proline is a structurally important amino acid, and it's conservation at this residue suggests it plays an important role in Pus4 function (Deber et al., 2010). We hypothesize the mutation from proline to serine disrupts the structure of Pus4A, rendering it nonfunctional. Additionally, we examined the promoter sequence

upstream of both alleles and found there were differences in sequence. One of these differences introduces an Aft2 binding site at residue -99 upstream of the B allele but not the A allele. (Appendix 3) (Teixeira et al., 2018). Additionally, there is a second binding site for Mot3 upstream of *PUS4A* that is not present in *PUS4B* (Appendix 3). Aft2p is an iron dependent transcriptional activator that has been shown to have a role in *C. albicans* morphogenesis and virulence (Liang et al., 2010). Mot3 is a transcriptional repressor that can play a role in filamentation (Khalaf and Zitomer, 2001). Together these results suggest a biological mechanism for decreased expression of *PUS4A*. The lack of a binding site for an activator and an extra site of a repressor likely turns off expression of *PUS4A* so that a non-functional protein is not produced.

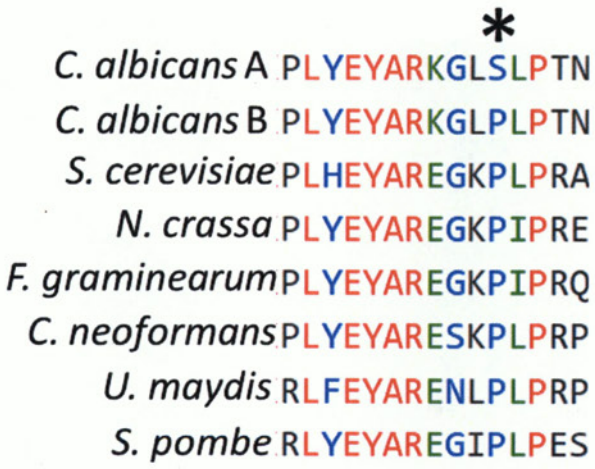

Figure 6: The mutation of proline to serine at residue 175 in *Candida albicans PUS4A*. The proline residue is conserved in almost all yeast species, including the species of note shown here.

Discussion

For much of our experiments, including the growth assay, we did not see a difference when the expression of one *PUS4* allele was changed. The lack of a difference suggests that Pus4 does not have a direct impact on growth rate, though this should be further explored. Though we did see a small difference in tolerance to oligomycin, it was small. This is likely because *Candida albicans* can survive without it's mitochondria, switching to fermentation for energy (Shingu-Vazquez and Traven, 2011). The ability to rely on fermentation likely prevents the mitochondrial dysfunction caused by oligomycin from having a larger impact on growth, which would make the change in tolerance more notable. Additionally we did not see a difference in virulence in *Galleria mellonella* in any of the strains. Because neither the increase or decrease in expression of *PUS4* results in a change in virulence it suggests that pathogenesis is not dependent upon precise levels of Pus4 in a cell.

Our sequencing results and further bioinformatic analysis help to explain the modest changes that we see in our phenotypic assays. The maltose promoter inserted upstream of *PUS4A* likely had little to no effect on expression because the allele is not expressed or expressed at very low levels normally. When we did see some small changes in phenotype, such as in the tolerance to oligomycin, they were as a result in changing the expression of *PUS4B*. This is because only the expression of *PUS4B* is changing. While our experiments helped to confirm some of the functions of Pus4 that were inferred from *Saccharomyces cerevisiae* more work should be done to verify functions of Pus4 in *Candida albicans*. This work should focus on *PUS4B*.

Combining our data from sequencing and the bioinformatic analysis allows us to propose a hypothesis for the elimination of transcription from *PUS4A*. We propose that the proline residue mutated to a serine residue, which rendered the protein nonfunctional. A nonfunctional Pus4A was no longer under pressure to be transcribed and this allowed a mutation in the binding site for Aft2p to occur without consequence, eliminating transcription of the faulty protein. In addition, one could posit that a nonfunction Pus4A could act a dominant negative and actually be harmful to the cells in which case turning off expression would be beneficial. This data also leads to a potential answer to the question of why *Candida albicans* maintains a diploid genome. Having two copies of the *PUS4* gene allowed the organism to compensate for mutations that lead to a loss of function. In this way the diploid genome can be used as a way for *Candida albicans* to compensate for mutations without a significant loss of function. This hypothesis can also help explain why haploid *Candida albicans* is less fit than diploid strains (Hickman et al., 2013), because they do not have the ability to compensate for mutations. Additionally, even though haploid *Candida albicans* can revert to diploid, it generates a homozygous diploid. If one allele of a gene, such as *PUS4A*, contained a mutation that resulted in a less or non-functional protein, a homozygous diploid might have two copies of that allele that leads to a negative effect, rendering it less fit.

In the future more work should be done to better classify the functions of Pus4B, particularly in conditions similar to infection. Furthermore, since Aft2p is an iron-dependent transcription factor, the influence of different iron concentrations on the expression of *PUS4B* and the function of Pus4B should be explored. Additionally, the targets of Pus4B should be determined, particularly the relatively unknown mRNA

targets. Better understanding of the functions of Pus4B will help further the understanding of both the role of pseudouridylation in *Candida albicans* and the pressure to maintain a diploid genome.

References

Becker, H.F., Motorin, Y., Planta, R.J., and Grosjean, H. (1997). The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of Ψ55 in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Research *25*, 4493-4499.

Brennan, M., Thomas, D.Y., Whiteway, M., and Kavanagh, K. (2002). Correlation between virulence of Candida albicans mutants in mice and Galleria mellonella larvae. FEMS Immunology & Medical Microbiology *34*, 153-157.

CDC (2013). Antibiotic Resistance Threats in The United States, 2013 (Centers for Disease Control and Prevention), pp. 63-64.

Deber, C., M., Brodsky, B., and Rath, A. (2010). Proline Residues in Proteins. eLS.

Gerstein, A.C., Rosenberg, A., Hecht, I., and Berman, J. (2016). diskImageR: quantification of resistance and tolerance to antimicrobial drugs using disk diffusion assays. Microbiology *162*, 1059-1068.

Hickman, M.A., Zeng, G., Forche, A., Hirakawa, M.P., Abbey, D., Harrison, B.D., Wang, Y.-M., Su, C.-h., Bennett, R.J., Wang, Y., *et al.* (2013). The 'obligate diploid' Candida albicans forms mating-competent haploids. Nature *494*, 55.

Khalaf, R.A., and Zitomer, R.S. (2001). The DNA Binding Protein Rfg1 Is a Repressor of Filamentation in <em>Candida albicans</em&gt. Genetics *157*, 1503.

Kullberg, B.J., and Arendrup, M.C. (2015). Invasive Candidiasis. New England Journal of Medicine *373*, 1445-1456.

Liang, Y., Wei, D., Wang, H., Xu, N., Zhang, B., Xing, L., and Li, M. (2010). Role of Candida albicans Aft2p transcription factor in ferric reductase activity, morphogenesis and virulence. Microbiology *156*, 2912-2919.

Oh, J., Fung, E., Schlecht, U., Davis, R.W., Giaever, G., St Onge, R.P., Deutschbauer, A., and Nislow, C. (2010). Gene annotation and drug target discovery in Candida albicans with a tagged transposon mutant collection. PLoS Pathog *6*, e1001140.

Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., *et al.* (2014). Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. Cell *159*, 148-162.

Shingu-Vazquez, M., and Traven, A. (2011). Mitochondria and Fungal Pathogenesis: Drug Tolerance, Virulence, and Potential for Antifungal Therapy. Eukaryotic Cell *10*, 1376-1383.

Skrzypek, M.S., Binkley, J., Binkley, G., Miyasato, S.R., Simison, M., and Sherlock, G. (2017). The Candida Genome Database (CGD): incorporation of Assembly 22,

systematic identifiers and visualization of high throughput sequencing data. Nucleic Acids Res *45*, D592-D596.
Teixeira, M.C., Monteiro, P.T., Palma, M., Costa, C., Godinho, C.P., Pais, P., Cavalheiro, M., Antunes, M., Lemos, A., Pedreira, T.*, et al.* (2018). YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in Saccharomyces cerevisiae. Nucleic Acids Research *46*, D348-D353.
Zhao, Y., Dunker, W., Yu, Y.T., and Karijolich, J. (2018). The Role of Noncoding RNA Pseudouridylation in Nuclear Gene Expression Events. Front Bioeng Biotechnol *6*, 8.

Appendix 1: Oligos used in this study

Description	Allele	Sequence 5'-3'
Maltose promoter forward transformation guide	PUS4A	gatgtcgcactaccacacgcaatagaaaaatagaatcctg**gtaccgaattcgagctcgtt**
Maltose promoter reverse transformation guide	PUS4A	**accactgcaaaaagttgggg**atgtaatgcctgagggtttgtatactgcaaatactccgtt
Maltose promoter sequencing primer	Both	**atgagtgatgatgtcgcact**
Maltose promoter forward transformation guide	PUS4B	gatttcgcactaccacacccaatagaaaaatagaaacctg**gtaccgaattcgagctcgtt**
Maltose promoter reverse transformation guide	PUS4B	**accactgcaaaaagttgggg**atgtaatgcctgagggtttgtatactgcaaatactccgtt

Sequences that are complementary to the maltose promoter are shown in bold, while sequences that are complementary to *Candida albicans* are shown in regular font.

Appendix 2: Strains used in this study

Strain ID	Alternate Strain ID	Organism	Description
CE1	DAB 690	*Candida albicans*	SC5314, Wildtype
CE 16	DAB 801	*Candida albicans*	Maltose inducible promoter upstream of *PUS4B*
CE17	DAB 802	*Candida albicans*	Maltose inducible promoter upstream of *PUS4A*

Appendix 3: Transcription factors upstream of *PUS4A* and *PUS4B*

Transcription Factor Binding Site	Consensus Sequence	BP Upstream of *PUS4 A*
Ash1	YTGAT	349
Ash1	YTGAT	336
Ash1	YTGAT	313
Ash1	YTGAT	196
Ash1	YTGAT	145
Azf1	AAAAGAAA	380
Azf1	AAAAGAAA	325
Basf1,Gcn4	TGACTC	283
Fkh1,Fkh2	RYMAAYA	308
Fkh1,Fkh2	RYMAAYA	272
Fkh1,Fkh2	RYMAAYA	44
Gcn4	TGATTCA	498
Gcn4	TGATTMA	284
Gcn4	RRTGACTC	281
Gcr1	CWTCC	461
Gcr1	CWTCC	155
Mac1	TTTGCKCR	467
Mot3	CAGGYA	468
Mot3	**TMGGAA**	**416**
Mot3	AAGGWT	174
Rtg1, Rtg3	GTCAC	287
Stb5	CGGNS	290
Stp2	CNCACCNG	425
Xbp1	CTCGA	458
Yap1, Cad1, Yap3, Cin5, Yap5	TTACTAA	393
Yap1	TKACAAA	385
Yap1	TTACTCA	215

Transcription Factor Binding Site	Consensus Sequence	BP Upstream of *PUS4 B*
Aft2, Aft1	**YRCACCCR**	**99**
Ash1	YTGAT	350
Ash1	YTGAT	337
Ash1	YTGAT	314
Ash1	YTGAT	195
Ash1	YTGAT	144
Azf1	AAAAGAAA	381
Azf1	AAAAGAAA	326
Basf1,Gcn4	TGACTC	284
Fkh1,Fkh2	RYMAAYA	309
Fkh1,Fkh2	RYMAAYA	272
Fkh1,Fkh2	RYMAAYA	44
Gcn4	TGATTCA	498
Gcn4	TGATTMA	284
Gcn4	RRTGACTC	282
Gcr1	CWTCC	462
Gcr1	CWTCC	154
Mac1	TTTGCKCR	468
Mot3	CAGGYA	469
Mot3	AAGGWT	173
Rtg1, Rtg3	GTCAC	288
Stb5	CGGNS	291
Stp2	CNCACCNG	426
Xbp1	CTCGA	459
Yap1, Cad1, Yap3, Cin5, Yap5	TTACTAA	394
Yap1	TKACAAA	386
Yap1	TTACTCA	214